

DATA ALBUMS: An EVENT DRIVEN SEARCH, AGGREGATION and CURATION TOOL FOR EARTH SCIENCE

Rahul Ramachandran¹, Ajinkya Kulkarni², Manil Maskey², Rohan Bakare², Sabin Basyal², Xiang Li² and Shannon Flynn²

¹NASA/MSFC, ²University of Alabama in Huntsville

ABSTRACT

Approaches used in Earth science research such as case study analysis and climatology studies involve gathering, discovering and gathering diverse data sets and information to support the research goals. To gather relevant data and information for case studies and climatology analysis is both tedious and time consuming. Current earth science data systems are designed with the assumption that researchers access data primarily by instrument or geophysical parameter. In cases where researchers are interested in studying a significant event, they have to manually assemble a variety of datasets relevant to it by searching the different distributed data systems. This paper presents a specialized search, aggregation and curation tool for Earth science to address these challenges. The search tool automatically creates curated "Data Albums", aggregated collections of information related to a specific event, containing links to relevant data files (granules) from different instruments, tools and services for visualization and analysis and information about the event contained in news reports, images or videos to supplement research analysis. Curation in the tool is driven via an ontology based relevancy ranking algorithm to filter out non-relevant information and data.

INFORMATION SEEKING MODELS

STANDARD MODEL

- Identifying the problem (task)
- Articulating the information need
- Formulating the result
- Evaluating the result



Drawback - assumption that the task remains the same

DYNAMIC MODEL (Bates, 1989)

- User's needs evolve during the process as they interact with the information
- Analogous to animal foraging for food - learning as they move from one food resource to another
- Information foraging



Interaction with information leads to new unanticipated goals - "discovery of latent needs"

M. J. Bates, "The Design of Browsing and Browsing Techniques for the Online Search Interface," Online Rev., vol. 11, no. 5, pp. 407-431, 1989

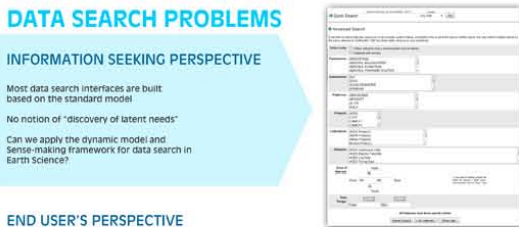
SENSE-MAKING - Behavior generally applied to intelligence analysis and other knowledge workers related to information seeking and use (Bates, 1983)

Framework for sense-making for intelligence analyst (Piroli and Card, 2005)

- Shoebox - gathering relevant documents into a single collection
- Evidence file - shoebox is curated to be filtered further
- Schema - building a model on how all the information fits

B. Davis, "An overview of sense-making research: Concepts, methods and results," in Annual Meeting of the International Communication Association, Dallas, TX, 1983

P. Piroli and S. Card, "The Sensemaking Process and Leverage Points for Analyst Technology Identified Through Cognitive Task Analysis," 3333 Cognitive Hill Road 2, A National Model of Analyst Sense-making, in Proceedings of the 2005 International Conference on Intelligence Analysis, Moscow, VA, 2005.



ONTOLOGY-BASED RELEVANCY RANKING SERVICE

Designed as a general service that can be customized for specific applications. Utilizes an algorithm that combines ontology based and traditional statistical score to estimate relevancy of a resource (Bouramoul and Khodja, 2012; Shamsfard et al., 2006)

A. Bouramoul and M. Khodja, "An ontology based approach for semantic ranking of the web search engines results," in 2012 International Conference on Multimedia Computing and Systems (MMCS), 2012.

M. Shamsfard, A. Nematollahi, and S. Moti, "OBRS: An Ontology Based System for Ranking Documents," International Journal of Computer Science, vol. 1, no. 3, pp. 225-231, 2006.

Relevance Score: Sum of the score of each matched concept obtained by multiplying ontology-based score and the statistical score

$$S_r = \sum_{i=1}^n I(c_i) \times I(f_i) \times I(d_i, D)$$

DATA ALBUMS CONCEPT

Aggregated results are presented with visual interactive interfaces to support:

- "Discovery of latent needs"
- Dual coding theory - information is best grasped when presented in two modalities

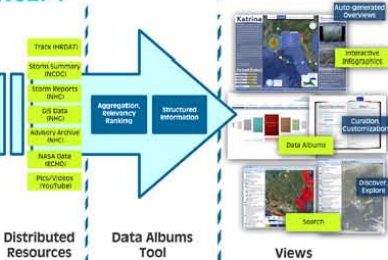
Integrates the "Shoebox" and the "Evidence File" sensemaking concepts

Discovery and aggregation tool that automatically generates "Data Albums"

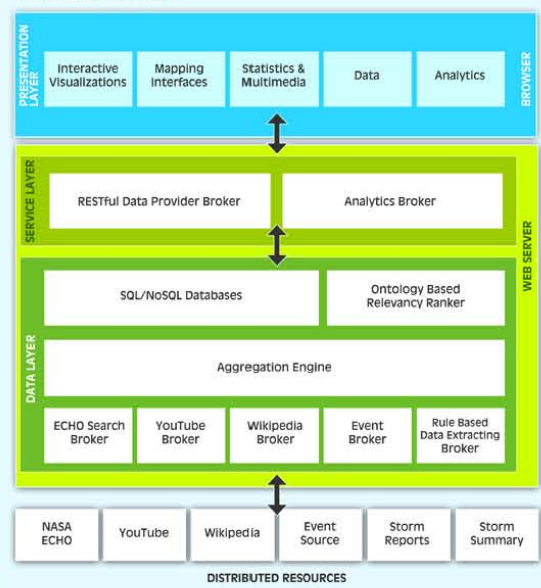
- Is a compiled collection of data, information, and tools around an event or a theme to support scientific research

Different distributed resources are semantically curated using an application ontology

- Used both for query expansion and for relevancy ranking text mining algorithm



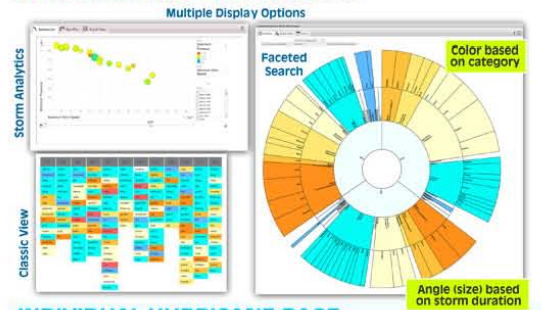
ARCHITECTURE



TWO SCIENCE FOCUS AREAS

- Hurricane Case Studies at GHRC**
 - Focus on hurricane events, especially those observed during past NASA field campaigns
 - Aggregate data from different airborne sensors with satellite observations and supplemental information available online
- Severe Storm Case Studies at NASA's SPoRT**
 - Focus on challenging weather phenomena such as convective thunderstorms
 - Help automate the selection of weather events and other information needed for evaluating SPoRT's mesoscale configuration of the Weather Research and Forecasting (WRF) model for convective cases

OVERVIEW PAGE - HURRICANES



INDIVIDUAL HURRICANE PAGE

Relevant imagery from GIBS for Context

Different Data Collections Relevant for this Event



SEVERE STORMS



CONTACT
rahul.ramachandran@nasa.gov
<http://innovations.itsc.uah.edu/dataalbums/>

ACKNOWLEDGEMENT
 This work was conducted as part of the NASA ACCESS Grant #75S161.